#24

DECLARATION

I Harutoshi Suzuki, of 3-6-5 Fujigaoka, Fujisawa-shi, Kanagawa-ken, 251-0004 Japan, declare that I am a Patent Agent and conversant with the Japanese and English languages and that the accompanying translation, which was prepared by me, is a true translation of Japanese Patent Application No. 9-296050.

Signed this 11th day of August 2003

Harutoshi Suzuki

PATENT OFFICE
JAPANESE GOVERNMENT

This is to certify that the annexed is a true copy of the following application as filed with this Office.

Date of Application: October 28, 1997

Application Number: Tokuganhei 9-296050

Applicant(s): YAMAHA CORPORATION

October 2, 1998

Commissioner, Patent Office

Takeshi ISAYAMA

Proof No. 10-3079253

[Document] Patent Application

[Docket No.] C26812

[Filing date] October 28, 1997

[Addressee] Commissioner of Patent Office

[I.P.C.] G10L 5/00

[Title] Sound Converter

[Number of Claims] 9

[Inventor]

[Residence] c/o YAMAHA CORPORATION

10-1 Nakazawa-cho, Hamamatsu-shi, Shizuoka-ken, Japan

[Name] Yasuo YOSHIOKA

[Applicant]

[Identification Code] 000004075

[Name] YAMAHA CORPORATION

[Agent]

[Identification Code] 100098084

[Patent Agent]

[Name] Kenji KAWASAKI

[Designated Agent]

[Identification Code] 100104798

[Patent Agent]

[Name]	Tomoteru YAMASHITA	
[Indication of Fee]		
[Account No.]	038265	
[Amount]	21000	
[List of Items]		
[Item]	Specification	1
[Item]	Drawings	1
[Item]	Abstract	1
[Power of attorney]	9600966	
[Proof]	YES	

[Document name] Specification

[Title of the invention] SOUND CONVERTER

[Scope of claims]

[Claim 1] A sound converter comprising:

sinusoidal wave component extracting means for extracting a plurality of sinusoidal wave components from an input sound signal;

reference pitch information storing means for storing reference sound pitch information;

frequency adjusting means for adjusting the frequency of said sinusoidal wave components on the basis of pitch information read out from said reference pitch information storing means; and

synthetic waveform generating means for generating a synthetic waveform by synthesizing each of said sinusoidal wave components after said frequency adjustment by said frequency adjusting means.

[Claim 2] A sound converter comprising:

sinusoidal wave component extracting means for extracting a plurality of sinusoidal wave components from an input sound signal;

amplitude information storing means for storing amplitude information indicating the amplitude of a plurality of sinusoidal wave components extracted from a reference sound;

amplitude adjusting means for adjusting the amplitude of said sinusoidal wave components on the basis of amplitude information read out from said amplitude information storing means; and

synthetic waveform generating means for generating a synthetic waveform by synthesizing each of said sinusoidal wave components after said amplitude adjustment by said amplitude adjusting means.

[Claim 3] A sound converter comprising:

sinusoidal wave component extracting means for extracting a plurality of sinusoidal wave components from an input sound signal;

reference pitch information storing means for storing reference sound pitch information;

amplitude information storing means for storing amplitude information indicating the amplitude of a plurality of sinusoidal wave components extracted from said reference sound;

amplitude adjusting means for adjusting the amplitude of said sinusoidal wave components on the basis of amplitude information read out from said amplitude information storing means;

frequency adjusting means for adjusting the frequency of said sinusoidal wave components on the basis of pitch information read out from said reference pitch information storing means; and

synthetic waveform generating means for generating a synthetic waveform by synthesizing each of said sinusoidal wave components after said frequency adjustment and said amplitude adjustment by said frequency adjusting means and said amplitude adjusting means.

[Claim 4] The sound converter according to either of claims 1 or 3, wherein said frequency adjusting means varies the degree to which said pitch information relating to said sinusoidal wave components is reflected, in accordance with a prescribed parameter.

[Claim 5] The sound converter according to any one of claims 1, 3, or 4, wherein said reference pitch storing means stores musical pitch, which changes in musical scale units, and a fluctuation component indicating pitch fluctuation of said musical pitch, and said frequency adjusting means adjusts the frequency of said sinusoidal wave components on the basis of both said musical pitch and said fluctuation component.

[Claim 6] The sound converter according to either of claims 2 or 3, wherein said amplitude adjusting means varies the degree to which said amplitude information relating to said sinusoidal wave components is reflected, in accordance with a prescribed parameter.

[Claim 7] The sound converter according to any one of claims 1 to 6, further comprising volume information storing means for storing volume information indicating volume changes in said reference sound; and volume adjusting means for adjusting the volume of said synthetic waveform on the basis of volume information read out from said volume information storing means.

[Claim 8] The sound converter according to any one of claims 1 to 7, further comprising pitch determining means for determining whether or not a pitch is present in said input sound signal; and switching means for outputting said input sound signal instead of said synthetic waveform, when said pitch determining means determines that the pitch is not present.

[Claim 9] The sound converter according to any one of claims 1 to 8, further comprising residual component extracting means for determining residual components of the sinusoidal wave components extracted by said sinusoidal wave component extracting means and said input sound signal; and adding means for adding the residual components extracted by said residual component extracting means to said synthetic waveform.

[Detailed description of the invention]

[0001]

[Technical field of the invention]

The present invention relates to a sound converter which causes a processed sound to imitate a further sound forming a target.

[0002]

[Prior art]

Various sound converters which change the frequency characteristics, or the like, of an input sound and then output the sound, have been disclosed. For example, there exist *karaoke* devices which change the pitch of the singing voice of a singer to convert a male voice to a female voice, or vice versa (for

example, Publication of a Translation of an International Application No. Hei. 8-508581).

[0003]

[Problem to be solved by the invention]

However, in a conventional sound converter, although the voice is converted, this has simply involved changing the voice characteristics. Therefore, it has not been possible to convert the sound such that it approximates someone's voice, for example.

Moreover, it would be very amusing if a *karaoke* machine were provided with an imitating function whereby not only the voice characteristics, but also the manner of singing, could be made to sound like a particular singer. However, in conventional sound converters, processing of this kind has not been possible.

[0004]

The present invention was devised with the foregoing in view, an object thereof being to provide a sound converter which is capable of making voice characteristics imitate a target voice.

It is a further object of the present invention to provide a sound converter which is capable of making an input voice of a singer imitate the singing manner of a desired singer.

[0005]

[Means for solving the problem]

In order to resolve the aforementioned problems, a sound converter according to claim 1 comprises: sinusoidal wave component extracting means for extracting a plurality of sinusoidal wave components from an input sound signal; reference pitch information storing means for storing reference sound pitch information; frequency adjusting means for adjusting the frequency of the sinusoidal wave components on the basis of pitch information read out from the reference pitch information storing means; and synthetic waveform generating means for a generating synthetic waveform by synthesizing each of the

sinusoidal wave components after frequency adjustment thereof by the frequency adjusting means.

[0006]

The sound converter according to claim 2 comprises: sinusoidal wave component extracting means for extracting a plurality of sinusoidal wave components from an input sound signal; amplitude information storing means for storing amplitude information indicating the amplitude of a plurality of sinusoidal wave components extracted from a reference sound; amplitude adjusting means for adjusting the amplitude of the sinusoidal wave components on the basis of amplitude information read out from the amplitude information storing means; and synthetic waveform generating means for generating a synthetic waveform by synthesizing each of the sinusoidal wave components after amplitude adjustment thereof by the amplitude adjusting means.

[0007]

The sound converter according to claim 3 comprises: sinusoidal wave component extracting means for extracting a plurality of sinusoidal wave components from an input sound signal; reference pitch information storing means for storing reference sound pitch information; amplitude information storing means for storing amplitude information indicating the amplitude of a plurality of sinusoidal wave components extracted from the reference sound; amplitude adjusting means for adjusting the amplitude of the sinusoidal wave components on the basis of amplitude information read out from the amplitude information storing means; frequency adjusting means for adjusting the frequency of the sinusoidal wave components on the basis of pitch information read out from the reference pitch information storing means; and synthetic waveform generating means for generating a synthetic waveform by synthesizing each of the sinusoidal wave components after frequency adjustment and amplitude adjustment thereof by the frequency adjusting means and the amplitude adjusting means.

[0008]

The sound converter according to claim 4 is a sound converter according to either of claims 1 or 3, wherein the frequency adjusting means varies the degree to which the pitch information relating to the sinusoidal wave components is reflected, in accordance with a prescribed parameter.

[0009]

The sound converter according to claim 5 is a sound converter according to any one of claims 1, 3, or 4, wherein the reference pitch storing means stores musical pitch, which changes in musical scale units, and a fluctuation component indicating pitch fluctuation of the musical pitch, and the frequency adjusting means adjusts the frequency of the sinusoidal wave components on the basis of both the musical pitch and the fluctuation component.

[0010]

The sound converter according to claim 6 is a sound converter according to either of claims 2 or 3, wherein the amplitude adjusting means varies the degree to which the amplitude information relating to the sinusoidal wave components is reflected, in accordance with a prescribed parameter.

[0011]

The sound converter according to claim 7 is a sound converter according to any one of claims 1 to 6, further comprising: volume information storing means for storing volume information indicating volume changes in the reference sound; and volume adjusting means for adjusting the volume of the synthetic waveform on the basis of volume information read out from the volume information storing means.

[0012]

The sound converter according to claim 8 is a sound converter according to any one of claims 1 to 7, further comprising: pitch determining means for determining whether or not a pitch is present in the input sound signal; and switching means for outputting the input sound signal instead of the synthetic

waveform, when the pitch determining means determines that a pitch is not present.

[0013]

The sound converter according to claim 9 is a sound converter according to any one of claims 1 to 8, further comprising: residual component extracting means for determining residual components of the sinusoidal wave components extracted by the sinusoidal wave component extracting means and the input sound signal; and adding means for adding the residual components extracted by the residual component extracting means to the synthetic waveform.

[0014]

[Embodiments of the invention]

1. Basic structure of the first embodiment

Next, an embodiment of the present invention is described. Fig. 1 is a block diagram showing the composition of an embodiment of the present invention. This embodiment relates to a case where a sound converter according to the present invention is applied to a *karaoke* machine, thereby constituting a *karaoke* machine whereby imitations can be performed.

[0015]

Firstly, the principles of this embodiment are described. Initially, a song by the person who is to be imitated is analyzed and the pitch thereof and the amplitude of the sinusoidal wave components therein are recorded. Sinusoidal wave components are then extracted from the current singer's voice, and the pitch and the amplitude of the sinusoidal wave components in the voice being imitated are used to affect these sinusoidal wave components. The affected sinusoidal wave components are synthesized to form a synthetic waveform, which is amplified and output. Moreover, the degree to which the wave components are affected can be adjusted by a prescribed parameter.

By means of the aforementioned processing, a sound waveform which reflects the voice characteristics and singing manner of the person to be imitated

is formed and this waveform is output whilst a karaoke performance is conducted.

[0016]

2. Detailed structure of the first embodiment

In Fig. 1, numeral 1 is a microphone, which gathers the singer's voice and outputs a voice signal Sv. This voice signal Sv is then analyzed by a high-speed Fourier transform section 2, and the frequency spectrum thereof is detected. The processing implemented by the high-speed Fourier transform section is carried out in prescribed frame units, so a frequency spectrum is created successively for each frame. Fig. 2 shows the relationship between a voice signal Sv and frames thereof. Symbol FL denotes a frame, and in this embodiment, each frame FL is set such that it overlaps partially with the previous frame FL.

[0017]

Numeral 3 denotes a peak detecting section for detecting peaks in a frequency spectrum. For example, the peak values marked by the X symbols are detected in the frequency spectrum illustrated in Fig. 3. A set of such peak values is output for each frame in the form of frequency value and amplitude value co-ordinates, such as (F0,A0), (F1,A1), (F2,A2), ... (FN,AN). Fig. 2 gives a schematic view of sets of peak values for each frame.

Next, a peak continuation section 4 determines links with the previous and subsequent frames for the set of peak values for each frame output by the peak detecting section 3. Peak values considered to form a link are subjected to link processing, such that a data series is created. Here, the link processing is described with reference to Fig. 4.

The peak values shown in section (A) of Fig. 4 were detected in the previous frame, and the peak values shown in section (B) of Fig.4 were detected in the subsequent frame. In this case, the peak continuation section 4 investigates whether peak values corresponding to each of the peak values detected in the preceding frame, (F0,A0), (F1,A1), (F2,A2), ... , (FN,AN),

are also detected in the current frame. It determines whether the corresponding peak values are present according to whether or not a peak is currently detected within a prescribed range about the frequencies of the peak values detected in the preceding frame. In the example in Fig. 4, peak values corresponding to (F0,A0), (F1,A1), (F2,A2), ... are discovered, but a peak value corresponding to (FK,AK) is not observed.

[0018]

If the peak continuation section 4 discovers corresponding peak values, then they are coupled in time series order and output as a data series of sets. If it does not find a corresponding peak value, then the peak value is overwritten by data indicating that there is no corresponding peak for that frame. Fig. 5 shows one example of change in peak frequencies F0 and F1. Change of this kind also occurs in the amplitudes A0, A1, A2, ... In this case, the data series output by the peak continuation section 4 contains scattered values output at alternate frame intervals.

The peak values output by the peak continuation section 4 are called deterministic components thereafter. This signifies that they are components of the original signal (in other words, voice signal Sv) which can be rewritten definitely as sinusoidal wave elements. Each of the rewritten sinusoidal waves (precisely, the amplitude and frequency which are the parameters of the sinusoidal wave) are called partial components.

[0019]

Next, the interpolating and waveform generating section 5 carries out interpolation processing with respect to the deterministic components output from the peak continuation section 4, and it generates a waveform based on the deterministic components after interpolation. In this case, the interpolation is carried out at intervals corresponding to the sampling rate (for example, 44.1 kHz) of the final output signal (signal immediately prior to input to the amplifier 50 described hereinafter). The solid lines shown on Fig. 5 illustrate a

case where interpolation processing is carried out with respect to peak values F0 and F1.

Here, Fig. 7 shows the composition of the interpolating and waveform generating section 5. The elements 5a, 5a, ... shown in this diagram are respective partial waveform generating sections, which generate sinusoidal waves corresponding to the specified frequency value and amplitude value. Here, the partial components (F0,A0), (F1,A1), (F2,F3), ... in the present embodiment change from moment to moment in accordance with the respective interpolations, so the waveforms output from the partial waveform generating sections 5a, 5a, ... follows these changes. In other words, since partial components (F0,A0), (F1,A1), (F2,A2), ... are output successively by the peak continuation section 4 and are each subjected to interpolation, respectively, each of the partial waveform generating sections 5a, 5a, ... outputs a waveform whose frequency and amplitude fluctuates within a prescribed frequency range. The waveforms output by the respective partial waveform generating sections 5a, 5a, ... are added and synthesized at an adding section 5b. Therefore, the output signal from the interpolating and waveform generating section 5 has a waveform wherein the deterministic components have been extracted from the original signal (in other words, the voice signal Sv).

[0020]

Next, the deviation detecting section 6 shown in Fig. 1 calculates the deviation between the deterministic component waveform output by the interpolating and waveform generating section 5 and the voice signal Sv. Hereinafter, deviation components are called residual components Srd. The residual components comprise a large number of voiceless components contained in the sound. The aforementioned deterministic components, on the other hand, correspond to voiced components. When imitating someone's voice, the voiced sound only is processed and there is no particular need to process the voiceless sound. Therefore, in this embodiment, sound conversion processing is carried out only

with respect to the deterministic components corresponding to the voiced components.

Next, numeral 10 shown in Fig. 1 denotes a separating section, where the frequency values F0 – FN and amplitude values A0 – AN are separated from the data series output by the peak continuation section 4. The pitch detecting section 11 detects the pitch of each frame on the basis of frequency values supplied by the separating section 10. In the pitch detection process, a prescribed number of (for example, approximately three) frequency values are selected from the lowest of the frequency values output by the separating section 10, a prescribed weighting is applied to these frequency values, and the average thereof is calculated to give a pitch PS. Furthermore, for frames in which a pitch cannot be detected, the pitch detecting section 11 outputs a signal indicating that there is no pitch. A frame containing no pitch occurs in cases where the voice signal Sv in the frame is constituted almost entirely by voiceless sound and noise. In frames of this kind, since the frequency spectrum does not form a harmonic structure, it is determined that there is no pitch.

[0021]

Next, numeral 20 is a target information storing section wherein information relating to the object whose sound is to be imitated (hereinafter, called the target) is stored. The target information storing section 20 holds information on the target for separate songs. The target information comprises pitch information PTo containing the extracted musical pitch of the target sound, a pitch fluctuation component PTf, and deterministic amplitude components (same components as amplitude values A0, A1, A2, ... output by the separating section 10.) These information elements are stored respectively in a musical pitch storing section 21, a fluctuation pitch storing section 22 and a deterministic amplitude component storing section 23.

The target information storing section 20 is composed such that the respective items of information described above are read out in synchronism

with a *karaoke* performance. A *karaoke* performance is implemented in the performance section 27 illustrated in Fig. 1. Song data for use in *karaoke* is previously stored in the performance section 27, and song data selected by selecting means (omitted from diagram) is read out successively as the music proceeds, and supplied to an amplifier 50. In this case, the performance section 27 supplies a control signal Sc indicating the song title and the state of progress of the song to the target information storing section 20, which proceeds to read out the aforementioned information elements on the basis of this control signal.

[0022]

Next, the pitch information PTo read out from the musical pitch storing means 21 is mixed with the pitch PS in a ratio control section 30. This mixing is carried out on the basis of the following equation.

$(1.0 - \alpha)*PS + \alpha*PTo$ ------(1)

Here, α is a parameter which may take a value from zero to 1. The signal output from the ratio control section 30 is equal to pitch PS when $\alpha = 0$, and it is equal to pitch information PTo when $\alpha = 1$. Furthermore, parameter α is set to a desired value by means of an operator controlling a parameter setting section 25. The parameter setting section 25 can also be used to set the parameters β and γ, which are described hereinafter.

[0023]

Next, a pitch normalizing section 12 as illustrated in Fig. 1 divides each of the frequency values F0 – FN output from the separating section 10 by the pitch PS, thereby normalizing the frequency values. Each of the normalized frequency values F0/PS – FN/PS (dimensionless) is multiplied by the signal from the ratio control section by means of a multiplier 15, and the dimension thereof becomes frequency once again. In this case, it is determined from the value of parameter α whether the pitch of the singer inputting his or her voice

via the microphone 1 has a larger effect or whether the target pitch has a larger effect.

The ratio control section 31 multiplies the fluctuation component PTf output from the fluctuation pitch storing section 22 by the parameter β (where $0 \leq \beta \leq 1$), and outputs the result to a multiplier 14. In this case, the fluctuation component PTf indicates the divergence relating to the pitch information PTo in cent units. Therefore, the fluctuation component PTf is divided by 1200 (1 octave is 1200 cents) in the ratio control section 31, and calculation for finding the second power thereof is carried out, namely, the following calculation:

POW(2,(PTf*β / 1200)

The calculation results and the output signal from the multiplier 15 is multiplied by the multiplier 14. The output signal from the multiplier 14 is further multiplied by the output signal of a transposition control section 32 at a multiplier 17. The transposition control section 32 outputs values corresponding to the musical interval through which transposition is performed. The degree of transposition is set as desired. Normally, it is set to no transposition, or a change in octave units is specified. A change in octave units is specified in cases where there is an octave difference in the musical intervals being sung, for instance, where the target is male and the singer is female (or vice versa).

As described above, the target pitch and fluctuation component are appended to the frequency vales output from the pitch normalizing section 12, and if necessary, octave transposition is carried out, whereupon the signal is input to a mixer 40.

[0024]

Next, 13 illustrated in Fig. 1 is an amplitude detecting section, which detects the mean MS of the amplitude values A0, A1, A2, ... supplied by the separating section 10. In the amplitude normalizing section 16, the amplitudes values A0, A1, A2 are normalized by dividing them by this mean value. In the ratio control section 18, the deterministic amplitude components AT0, AT1,

AT2 ... (normalized) which are read out from the deterministic amplitude component storing section 23, are mixed with the aforementioned normalized amplitude values. The degree of mixing is determined by the parameter γ. If the deterministic amplitude components AT0, AT1, AT2, ... are represented by ATn (n = 1, 2, 3, ...), and the amplitude values output by the amplitude normalizing section 16 are represented by ASn' (n = 1, 2, 3, ...), then the operation of the ratio control section 18 can be expressed by the following calculation.

$(1 - \gamma)*ASn' + \gamma*ATn$

γ is a parameter set as appropriate in the parameter setting section 25, and it takes a value from zero to one. The larger the value of γ, the greater the effect of the target. Since the amplitude of the sinusoidal wave components in the voice signal determine voice characteristics, the voice becomes closer to the characteristics of the target, the larger the value of γ.

The output signal from the ratio control section 18 is multiplied by the mean value MS in a multiplier 19. In other words, it is converted from a normalized signal to a signal which represents the amplitude directly.

[0025]

Next, in the mixer 40, the amplitude values and the frequency values are combined. This combined signal comprises the deterministic components of the voice signal Sv of the singer, with the deterministic components of the target added thereto. Depending on the values of the parameters α, β and γ, 100% target-side deterministic components can be obtained.

These deterministic components (group of partial components which are sinusoidal waves) are supplied to the interpolating and waveform generating section 41. The interpolating and waveform generating section 41 is constituted similarly to the aforementioned interpolating and waveform generating section 5 (see Fig. 7). The interpolating and waveform generating section 41

interpolates the partial components contained in the deterministic components output from the mixer 40, and it generates partial waveforms on the basis of these respective partial components after interpolation and synthesizes these partial waveforms. The synthesized waveforms are added to the residual component Srd at an adder 42 and are then supplied via a switching section 43 to an amplifier 50. In frames where no pitch can be detected by the pitch detecting section 11, the switching section 43 supplies the amplifier 50 with the voice signal Sv of the singer instead of the synthesized signal output from the adder 42. This is because, since the aforementioned processing is not required for noise or voiceless sound, it is preferable to output the original signal directly.

[0026]

3. Operation of the first embodiment

Next, the operation of the embodiment having the foregoing composition is described. Firstly, when a song is specified, the song data for that song is read out by the performance section 27, and a musical sound signal is created on the basis of this data and supplied to the amplifier 50. The singer then starts to sing a song to this accompaniment, thereby causing a voice signal Sv to be output from the microphone 1. The deterministic components of this voice signal Sv are detected successively by the peak detecting section 3, frame by frame. For example, sampling results as illustrated in part (1) of Fig. 6 are obtained. Fig. 6 shows the signal obtained for a single frame. For each frame, links are created between partial components and these are separated by the separating section 10 and divided into frequency values and amplitude values, as illustrated in part (2) and (3) of Fig. 6. Furthermore, the frequency values are normalized by the pitch normalizing section 12 to give the values shown in part (4) of Fig. 6. The amplitude values are similarly normalized to give the values shown in part (5) of Fig. 6. The normalized amplitude values illustrated in part (5) of Fig. 6 are combined with normalized amplitude values for the target as shown in part (6)

to give amplitude values as shown in part (8). The ratio of this combination is determined by a parameter γ.

[0027]

Meanwhile, the frequency values shown in part (4) of Fig. 6 are combined with the target pitch information PTo and fluctuation component PTf to give the frequency values shown in part (7) of Fig. 6. The ratio of this combination is determined by parameters α and β. The frequency values and amplitude values shown in parts (7) and (8) of Fig. 6 are combined by the mixing section 40, thereby yielding new deterministic components as illustrated in part (9) of Fig. 6. These new deterministic components are formed into a synthetic waveform by the interpolating and waveform generating section 41, and this waveform is mixed with the residual components Srd and output to the amplifier 50.

As a result of the above, the singer's voice is output with the *karaoke* accompaniment, but the characteristics of the voice, the manner of singing, and the like, are significantly affected by the target. If the parameters α, β, γ are set to values of 1, the voice characteristics and singing manner of the target are adopted completely. In this way, singing which imitates the target precisely is output.

[0028]

4. Modifications

(1) As shown in Fig. 8, a normalized volume data storing section 60 for storing normalized volume data indicating changes in the volume of the target voice may also be provided. The normalized volume data read out from the normalized volume data storing section 60 is multiplied by a parameter k at a multiplier 61 and is then multiplied at a further multiplier with the synthesized waveform output from the switching means 43. By adopting the foregoing composition, it is even possible to imitate precisely the intonation of the target singing voice. The degree to which the intonation is imitated in this case is

determined by the value of the parameter. Therefore, the parameter k should be set according to the degree of imitation desired by the user.

[0029]

(2) In the present embodiment, the presence or absence of a pitch in a subject frame was determined by the pitch detecting section 11. However, detection of pitch presence is not limited to this, and may also be determined directly from the state of the voice signal Sv.

(3) Detection of sinusoidal wave components is not limited to the method used in the present embodiment. In short, it should be possible to detect sinusoidal waves contained in the voice signal.

(4) In the present embodiment, the target pitch and deterministic amplitude components were recorded. Alternatively, it is possible to record the actual voice of the target and then to read it out and extract the pitch and deterministic amplitude components by real-time processing. In other words, processing similar to that carried out on the voice of the singer in the present embodiment may also be applied to the voice of the target.

(5) In the present embodiment, both the musical pitch and the fluctuation component of the target were used in processing, but it is possible to use musical pitch alone. Moreover, it is also possible to create and use pitch data which combines the musical pitch and fluctuation component.

(6) In the present embodiment, both the frequency and amplitude of the deterministic components (set of sinusoidal wave components) of the singer's voice signal are converted, but it is also possible to convert either frequency or amplitude alone.

(7) In the present embodiment, a so-called oscillator system was adopted which uses an oscillating device for the interpolating and waveform generating section 5, 41. Besides this, it is also possible to use a reverse FFT, for example.

[0030]

[Advantages of the Invention]

As described above, according to the present invention, it is possible to convert a voice such that it imitates the voice characteristics and singing manner of a target.

[Brief description of the drawings]

[Fig. 1] This is a block diagram showing the composition of one embodiment of the present invention.

[Fig. 2] This is a diagram showing frame states according to the embodiment.

[Fig. 3] This is an illustrative diagram for describing the detection of frequency spectrum peaks according to the embodiment.

[Fig. 4] This is a diagram illustrating the linking of peak values for each frame according to the embodiment.

[Fig. 5] This is a diagram showing the state of change in frequency values according to the embodiment.

[Fig. 6] This is a graph showing the state of change of deterministic components during processing according to the embodiment.

[Fig. 7] This is a block diagram showing the composition of an interpolating and waveform generating section according to the embodiment.

[Fig. 8] This is a block diagram showing the composition of a modification of the embodiment.

[Description of references]

2---FAST FOURIERR TRANSFORM (SINOSOIDAL COMPONENT EXTRACTOR), 3---PEAK DETECTION (SINOSOIDAL COMPONENT EXTRACTOR), 4---PEAK CONTINUATION (SINOSOIDAL COMPONENT EXTRACTOR), 5---INTERPOLATING AND WAVEFORM GENERATING (RESIDUALCOMPONENT EXTRACTOR), 11---PITCH DETECTION (PITCH DETERMINTION), 12---PITCH NORMALIZATION, 13---AMPLITUDE DETECTION, 14,15AND 17---MULTIPLIER (FREQUENCY ADJUSTEMT MEANS), 16---AMPLITUDE NORMALIZATION, 20---TARGET SIGNAL STORING SECTION (REFERENCE PITCH STORAGE,

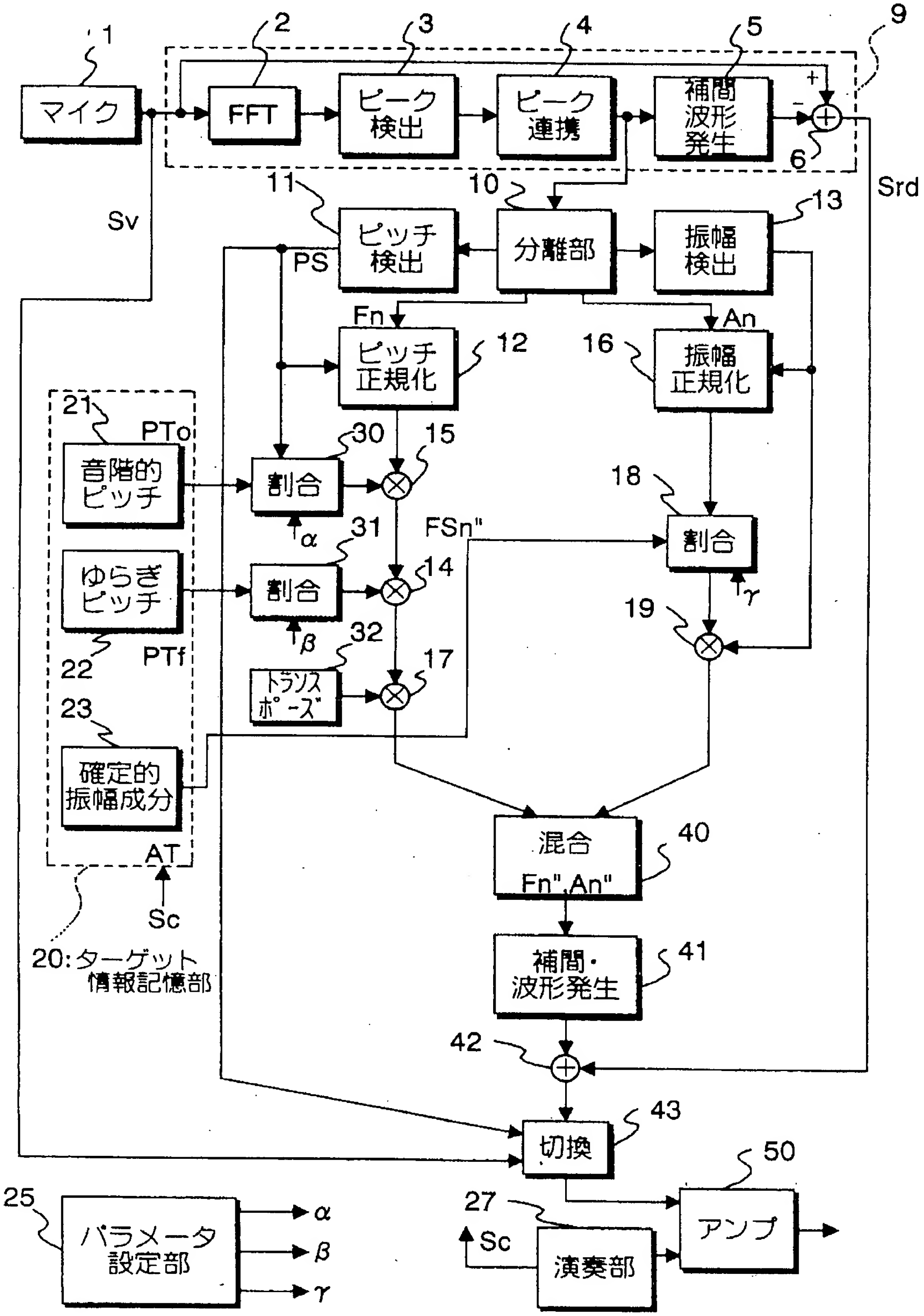
AMPLITUDE INFORMATION STORAGE), 25---PARAMETER SETTING SECTION, 30 AND 31---RATIO CONTROLLER (FREQUENCY ADJUASTMENT MEANS), 40---MIXING (SYNTHETIC WAVEFORM GENERATING MEANS), 41---INTERPORATION/WAVEFORM GENERATOR (SYNTHETIC WAVEFORM GENERTING MEANS), 42---ADDER, 43---SWITCH (SWITCHING MEANS), 60---NORMALIZED AMPLITUDE DATA STORAGE (AMPLITUDE INFORMATION STORING MEANS), 61 AND 62---MULTIPLIER (AMPLIFICTION ADJUSTEMNT MEANS)

FIG. 1

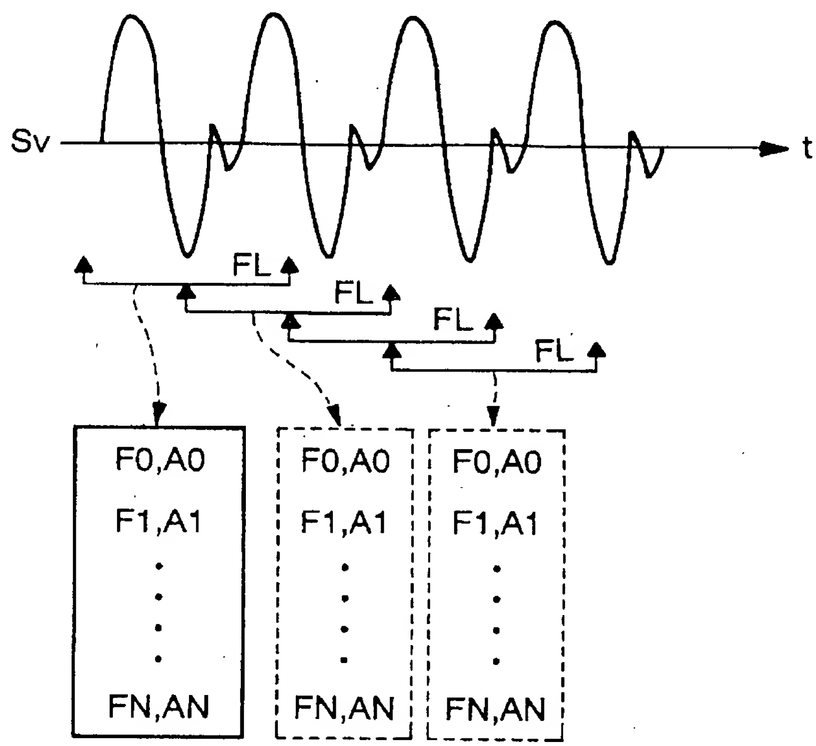
1	MICROPHONE
3	PEAK DETECTION
4	PICK CONTINUATION
5	INTERPOLATING AND WAVEFORM GENERATING
10	SEPARATING SECTION
11	PITCH DETECTION
12	PITCH NORMALIZATION
13	AMPLITUDE DETECTION
16	AMPLITUDE NORMALIZATION
18	AMPLITUDE NORMALIZATION
20	TARGET SIGNAL STORING SECTION
21	MUSICAL PITCH
22	PITCH FLUCTUATION
23	DETERMINISTIC AMPLITUDE COMPONENTS
25	PARAMETER SETTING SECTION
27	PERFORMANCE SECTION
30	RATIO
31	RATIO
32	TRANSPOSE
40	MIXING
41	INTERPOLATING AND WAVEFORM GENERATING
43	SWITCHING
50	AMPLIFIER

【書類名】 ~~図面~~ DRAWINGS

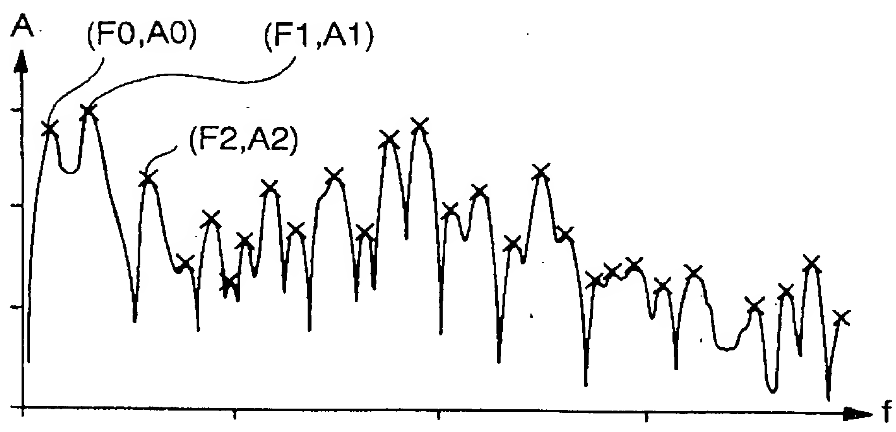
~~【図1】~~ FIG.1



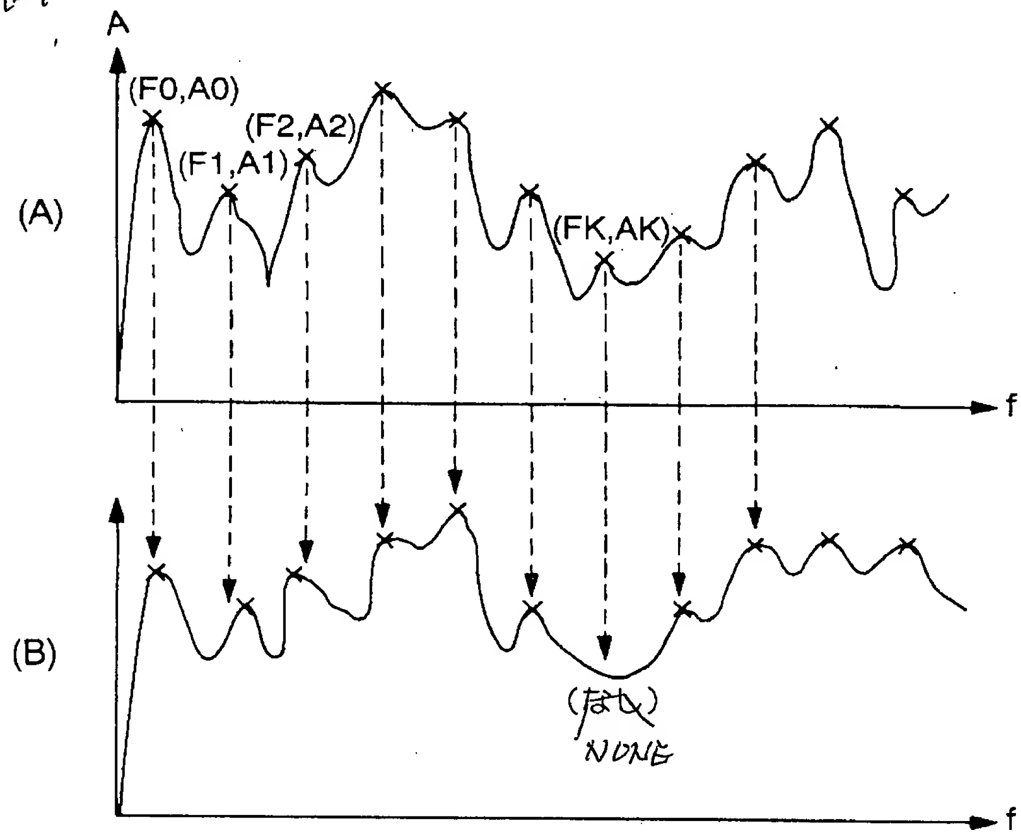
~~【図2】~~
FIG.2



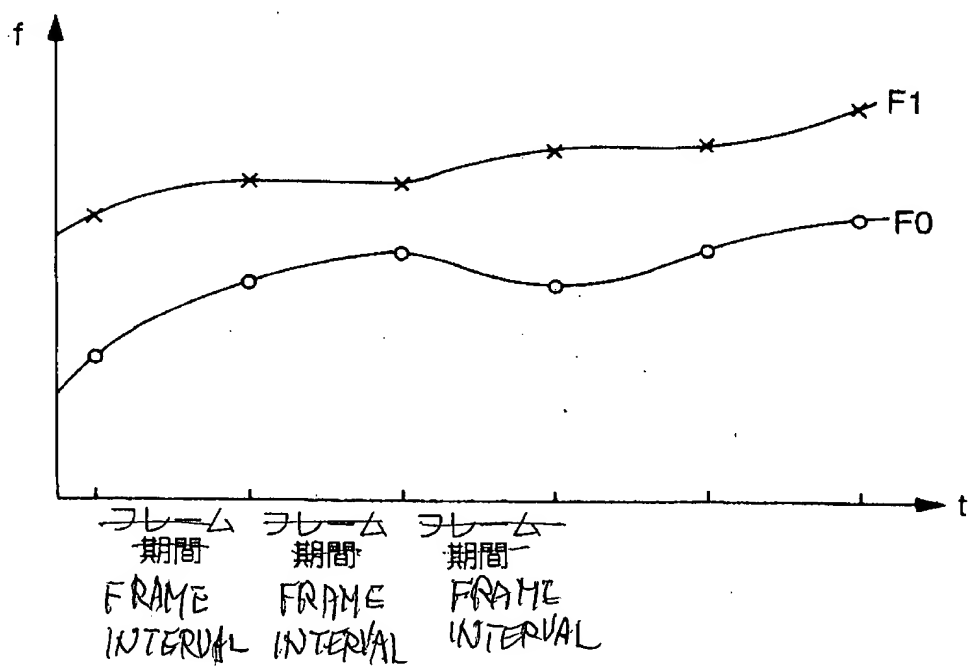
~~【図3】~~
FIG.3



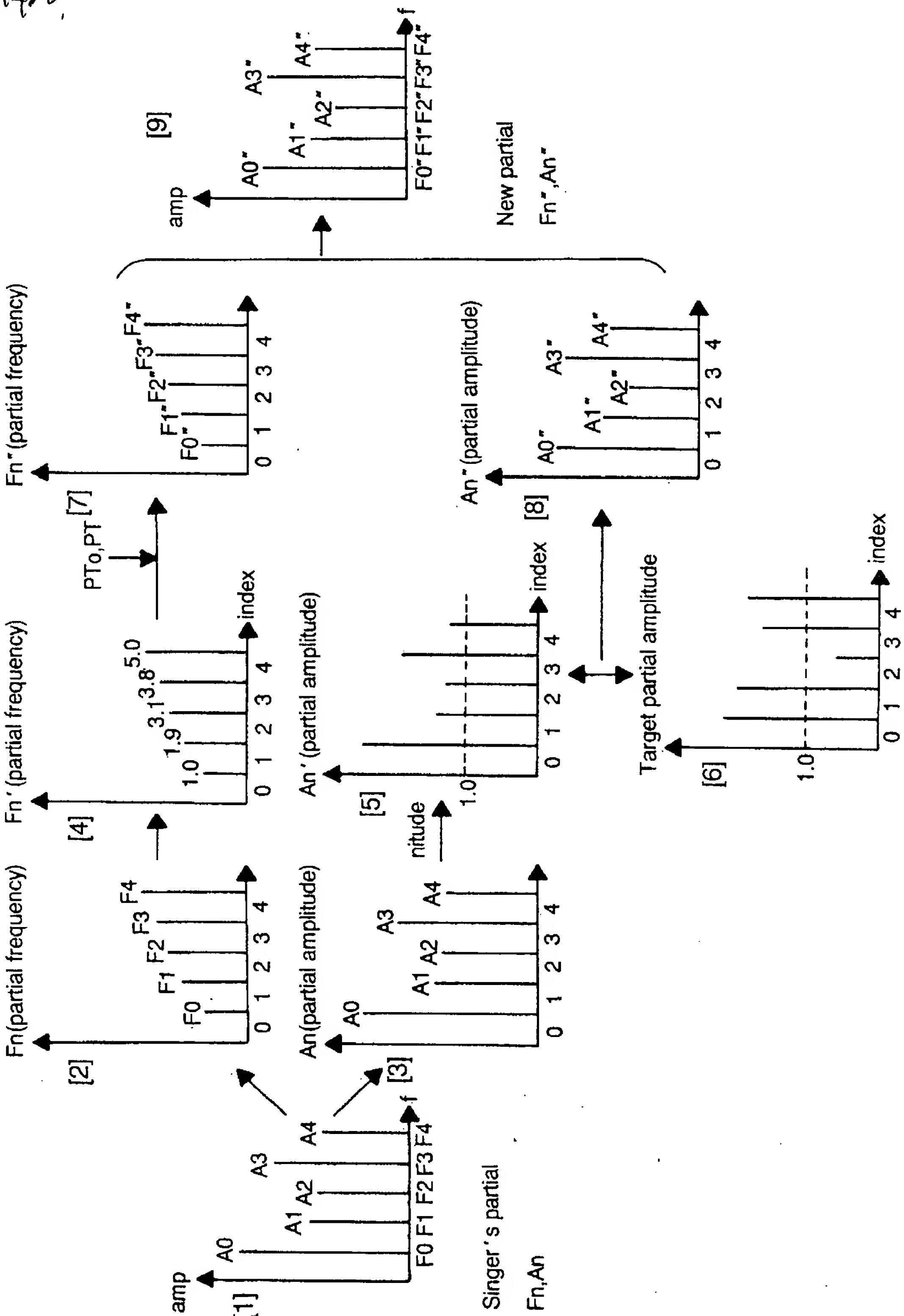
~~【図4】~~
FIG.4



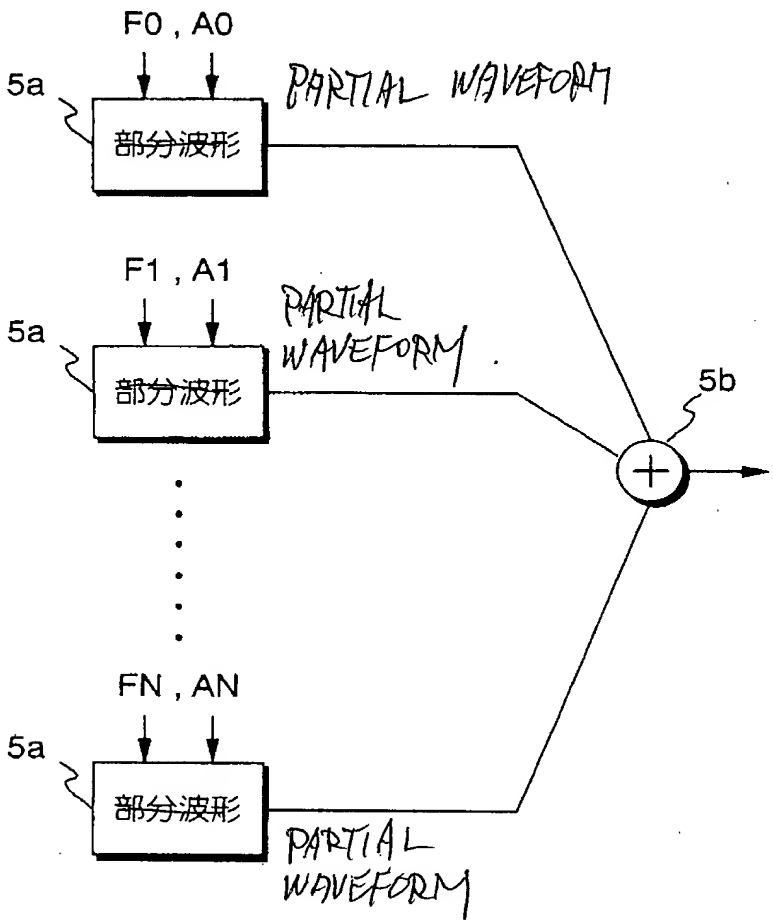
~~【図5】~~
FIG.5



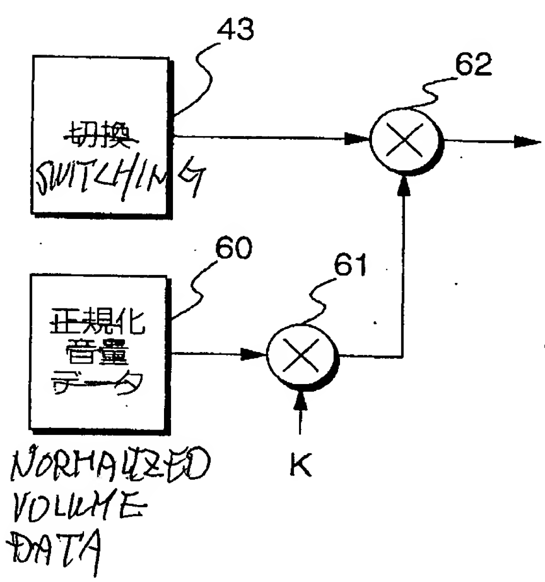
~~【図6】~~
FIG.6


[1]
amp
A0
A1
A2
A3
A4
F0 F1 F2 F3 F4
f
Singer's partial
Fn,An
[2]
Fn(partial frequency)
F0
F1
F2
F3
F4
0 1 2 3 4
[3]
An(partial amplitude)
[4]
Fn'(partial frequency)
1.0
1.9
3.1
3.8
5.0
index
[5]
An'(partial amplitude)
1.0
nitude
[6]
Target partial amplitude
[7]
PTo,PT
Fn"(partial frequency)
F0"
F1"
F2"
F3"
F4"
[8]
An"(partial amplitude)
A0"
A1"
A2"
A3"
A4"
[9]
New partial
Fn",An"
F0" F1" F2" F3" F4"

~~【図7】~~ FIG. 7


F0 , A0
5a
~~部分波形~~
PARTIAL WAVEFORM
F1 , A1
5a
~~部分波形~~
PARTIAL WAVEFORM
5b
FN , AN
5a
~~部分波形~~
PARTIAL WAVEFORM

【図8】


43
62
~~切換~~
SWITCHING
60
61
~~正規化音量データ~~
NORMALIZED VOLUME DATA
K

[Document] Amendment

[Filing date] May 12, 1998

[Addressee] Commissioner of Patent Office

[Indication of case] Tokuganhei 9-296050

[Requester]

[Relation to case] Applicant

[Identification Code] 000004075

[Name] YAMAHA CORPORATION

[Agent]

[Identification Code] 100098084

[Patent Agent]

[Name] Kenji KAWASAKI

[Amendment 1]

[Document] Patent Application

[Item] Inventors

[Method] Correction

[Contents]

[Inventor]

[Residence] c/o YAMAHA CORPORATION

10-1 Nakazawa-cho, Hamamatsu-shi, Shizuoka-ken, Japan

[Name] Yasuo YOSHIOKA

[Inventor]

[Residence] Biscaia 19, 2-2, 08440 Cardedeu, Barcelona, Spain

[Name] Xavier SERRA

[Amendment 2]

Hereafter omitted

DECLARATION

I, Xavier Serra, of Biscaia 19, 2-2, 08440 Cardedeu, Barcelona, Spain do hereby sincerely and solemnly declare that I invented the following invention with Yasuo Yoshioka:

Japanese Patent Application No. 9-296050
"Voice Changing Apparatus"

This 22nd day of April, 1998

By: ______________________
Xavier Serra

DEED OF ASSIGNMENT

Assignees,

Name : YAMAHA CORPORATION
Address: 10-1, Nakazawa-cho, Hamamatsu-shi, Shizuoka-ken, Japan

"Voice Changing Apparatus"

I, the undersigned assignor, Xavier Serra ,
do hereby affirm that I have assigned my rights to obtain a patent of the above-mentioned invention to the above-identified assignees.

This 2nd day of April, 1997

Assignor,

Address: Biscaia 19, 2-2, 08440 Cardedeu, Barcelona, Spain

Name: Xavier Serra

(Signature) ______________________